

# 实习总结

李伟豪

北京星天地信息科技有限公司

2024/05/28 - 2024/07/31

## 目录

<b>第一章 实习目标</b>	<b>1</b>
1.1 学习方向及目标	1
1.2 期望	1
<b>第二章 模型的搭建以及学习</b>	<b>2</b>
2.1 Yolo V5	2
2.1.1 模型简介	2
2.1.2 模型环境配置	2
2.1.3 模型训练	2
2.1.4 测试自己制作的数据集	3
2.1.5 实现原理	3
2.2 NeRF	5
2.2.1 模型简介	5
2.2.2 模型环境配置	5
2.2.3 模型训练	5
<b>致 谢</b>	<b>6</b>
<b>毕业设计小结</b>	<b>7</b>
<b>附 录</b>	<b>8</b>

## 第一章 实习目标

### 1.1 学习方向及目标

学习现在先进的计算机视觉（Computer Vision）以及图形学与 AI 结合的模型如 Neural Radiance Fields（神经辐射场，简称 NeRF），3D Gaussian Splatting（3D 高斯溅射，简称 3dgs），以及 YOLO（全称 You Only Look Once），同时理解各个模型的实现原理。

### 1.2 期望

- 完成搭建尽可能多的 AI 模型，配置其环境，并完成训练其预设数据库。
- 学习并理解各个模型的实现原理。
- 自己制作数据并通过基于 AI 的三维重建制作模型。

## 第二章 模型的搭建以及学习

### 2.1 Yolo V5

#### 2.1.1 模型简介

Yolo(You Only Look Once) 是一种单阶段目标检测算法，即仅需要“看”一次就可以识别出图片中物体的 class 类别和边界框。Yolov5 是由 Alexey Bochkovskiy 等人在 YOLO 系列算法的基础上进行改进和优化而开发的，使其性能与精度都得到了极大的提升。

#### 2.1.2 模型环境配置

- Python = 3.8.19
- torch = 2.3
- torchvision = 0.18.0
- gitpython = 2.40.1
- opencv-python = 4.9.0.80
- matplotlib = 3.7.5
- pandas = 2.0.3

模型从<https://github.com/ultralytics/yolov5.git> 克隆下来，然后在本地进行配置。

#### 2.1.3 模型训练

模型训练是在 `train.py` 文件中进行的，训练时可以同时输入 `data`<sup>1</sup>、`cfg`<sup>2</sup>和 `weight`<sup>3</sup>文件；或是 `epochs`<sup>4</sup>和 `batch size`<sup>5</sup>。

如下图：

---

<sup>1</sup>文件中包含了训练集和验证集的路径，以及所有的标注种类

<sup>2</sup>文件中包含了模型的参数，如学习率、batch size 等

<sup>3</sup>文件中包含了预训练模型的权重

<sup>4</sup>训练的次数

<sup>5</sup>每次训练的图片数量

```
python train.py --data coco.yaml --epochs 300 --weights '' --cfg yolov5n.yaml --batch-size 128
```

yolov5s	64
yolov5m	40
yolov5l	24
yolov5x	16

图 2-1 YOLOv5 目标检测模型的训练命令

#### 2.1.4 测试自己制作的数据集

从网上找寻了 30 张宝可梦的图片，然后通过 labelImg 工具<sup>6</sup>标注这些图片中比较常见的宝可梦，如皮卡丘、杰尼龟等，然后将这些图片和标注文件放入到一个文件夹中。最后通过 YOLOv5 模型进行训练，得到了一个可以识别这些宝可梦的模型，因为数据集比较小，所以模型的识别率不是很高。

训练结果示例：



图 2-2 YOLOv5 目标检测模型的训练结果

#### 2.1.5 实现原理

YOLOv5 首先会在输入端中将输入图片进行预处理，图像大小调整为模型所需的大小，进行归一化操作，及将像素值缩放到 0 到 1 之间。

然后将图片输入到 backbone 网络<sup>7</sup>中，backbone 网络会将图片的特征提取出来，然后将这些特征输入到 neck 网络<sup>8</sup>中，neck 网络会将不同层次的特征进行融合，然后将这些特征输入到 head 网络<sup>9</sup>中，head 网络会将图片中的物体进行预测，

<sup>6</sup>LabelImg 是一个开源的图像标注工具，用于创建图像数据集时标记图像中的对象边界框

<sup>7</sup>backbone 网络是一个特征提取网络，用于提取图片的特征

<sup>8</sup>neck 网络是一个特征融合网络，用于将不同层次的特征进行融合

<sup>9</sup>head 网络是一个预测网络，用于预测图片中的物体

得到物体的类别和边界框。

下图为 YOLOv5 的网络结构：<sup>10 11 12 13 14 15 16</sup>

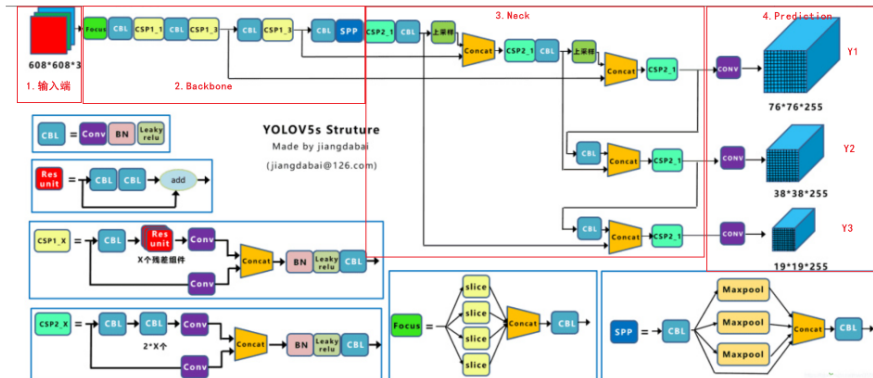


图 2-3 YOLOv5 的网络结构

<sup>10</sup>CPL: 由卷积 Conv+ 批量归一化 BN+ 激活函数 Leaky Relu 组成, 用于在特征提取的过程中增加网络的信息传递能力。CPL 通过在不同阶段的特征图之间进行部分连接, 使得网络可以更充分地利用低级和高级特征, 从而提高模型的性能。

<sup>11</sup>CSP1: 由 CBL 模块、Res unit 模块以及卷积层 Concat 组成。CSP1 是 CPL 的一种变体, 它在 CPL 的基础上引入了跨阶段的残差连接, 以进一步增强网络的信息传递能力。CSP1 可以有效地减少网络中的计算量, 提高模型的效率。

<sup>12</sup>CSP2: 不再使用 Res unit, 由卷积层 CBL 模块 Concat 组成。CSP2 引入了路径重排的技术, 以重新排列网络中的路径, 使得前向和后向传播之间的信息流更加平滑和高效。这种路径重排有助于减少网络中的计算量, 提高模型的速度和效率。CSP2 同样也进行了路径融合 (Path Fusion), 即在不同路径之间进行信息融合, 以使网络可以更充分地利用低级和高级特征。这种路径融合有助于提高模型对目标的检测和识别能力。CSP1 也具有类似的路径融合机制。

<sup>13</sup>SPP 是一种用于空间金字塔池化的技术, 它能够在不同尺度下提取图像的局部特征。SPP 在 YOLOv5 中用于提取特征图的空间信息, 从而使得模型可以更好地检测不同尺度和大小的目标。

<sup>14</sup>Focus: 对图像进行切片后再 Concat

<sup>15</sup>上采样: 利用元素复制扩充的方法使得特征图尺寸扩大, 例如线性插值。

<sup>16</sup>Concat: 张量拼接, 会扩充两个张量的维度, 实现多尺度特征融合。

## 2.2 NeRF

### 2.2.1 模型简介

NeRF (Neural Radiance Fields) 是一种用于三维重建的深度学习模型，它可以从二维图像中重建出高质量的三维场景。NeRF 的核心思想是利用神经网络来建模场景中每个空间点的辐射亮度和密度，从而实现对场景的准确重建。

### 2.2.2 模型环境配置

于 CUDA 10.2 基础下：

- Python = 3.7.1
- cudatoolkit = 10.0.130
- tensorflow = 1.14.0
- numpy = 1.21.5
- imageio = 2.19.3
- configargparse = 1.4

于 CUDA 12.x 基础下：

- Python = 3.7.1

模型从<https://github.com/bmild/nerf.git> 克隆下来，然后在本地进行配置。

### 2.2.3 模型训练

模型训练是在 `run_nerf.py` 文件中进行的，需要同时输入配置文件的相对路径，配置文件中包含了实验名称 (`exname`)，基础目录 (`basedir`)，数据集路径 (`datadir`)，数据集类型 (`dataset_type`) 采样参数、使用视角方向信息等模型训练和数据配置信息。

致谢

致 谢

致谢内容。

毕业设计小结



## 毕业设计小结

小结内容。

## 附录

### 附 录

附录内容。